

Evaluation Method for Automotive Stereo-Vision Systems

J. Morat^{*◇}, F. Devernay^{*}, J. Ibanez-Guzman[◇], and S. Cornou[◇]

Abstract—Safe vehicle guidance under human or computer control requires a thorough understanding of the traversed environment. Consequently if perception systems are to be introduced into mass market vehicles as part of driving assistance systems, their proper operation throughout the vehicle working life is needed. Onboard stereo-vision systems can provide rich information in terms of range, feature recognition, etc., hence the interest by car OEMs. System performance depends on multiple factors like light conditions, algorithms and the mechanical apparatus. Due to inaccuracies produced by changes in the system physical properties due to vibrations, misalignment of fixtures, etc. through the vehicle operational life a reduction in performance will occur. In this paper, an evaluation framework to estimate the performance of a vehicle onboard stereo-vision system in terms of 3D measurements and re-projection errors is presented. The approach considers changes that might occur in the system during the vehicle working life. It includes means to evaluate the self-calibration process often used to correct the effects of physical changes in the stereo-vision system. The results provide key information for the design and geometrical specification of automotive stereo-vision systems. As the potential physical changes in the geometric configuration of the camera-pair over the vehicle life time are difficult to predict, it was necessary to simulate them to generate families of errors that these might trigger on the system performance. The results of the observations and analysis are included; these should assist designers to define the constraints that needed for the layout of cameras as part of the design process.

I. INTRODUCTION

For several years passive and active sensors have been used for obstacle detection and vehicle guidance. Whilst active sensors such as RADAR and LADAR provide direct range measurements in a reliable manner, they suffer from low resolution, limited field of view and cost; nevertheless they could be very effective for different car applications and operate in complement to vision-based systems [1]. By contrast range estimations using stereo-vision systems offer dense images, ample field of views, at affordable cost. Their constraints, include the matching of features across images due to image noise [2], [3], [4], [5]. Another limitation originates in the estimation inaccuracy of the geometrical and optical parameters of the cameras due to the calibration process. Further, these parameters will vary when systems are mounted onboard vehicle platforms due to motion, vibration, shock and other external factors. Thus, there will be a performance degradation of such systems

over their working life. The successful deployment of stereo-vision systems in mass market vehicles will depend on the capability of ensuring the same performance despite physical changes that might occur. In this paper, a framework to quantify the performance of these systems is presented in terms of range estimations and re-projection errors. These are based on a theoretical formulation and extensive use of simulation techniques for validation purposes. The results are also used to determine the directions of the displacements and angles that might affect more the performance of a stereo-vision system. These are then formulated as a set of constraints to be taken into account by designers of stereo-vision systems. The remainder of this paper is organized as follows: Section II presents the camera parameters and the criteria used to quantify the system performance. Section III details the method developed to evaluate the effects of parameter deviations. It includes the associated results from simulation runs. Section IV portrays the self-calibration method adapted to automotive stereo-vision systems and the developed evaluation method. Section V concludes the paper and discusses results.

II. STEREO-VISION SYSTEMS FOR AUTOMOTIVE APPLICATIONS

A. Principle

A typical stereo-vision system uses a camera pair separated by a fixed baseline B to capture synchronised images from the same scene. The observation of the 3-D scene produces a parallax effect on the captured images. This difference is then used to recover depth information from the image pairs and to generate a disparity map. Figure 1 shows a typical configuration including the reference frames used in the analysis. C_l and C_r are the camera centers with reference frames $\mathcal{R}_l$ and $\mathcal{R}_r$, where $\mathcal{R}_v$ is the vehicle reference frame. A 3-D point Q observed by the cameras projects into the 2-D points q^l and q^r in the image planes. If the camera geometry is known, the 3-D coordinates of point Q can be recovered from the projections q^l and q^r in the image planes, a process known as 3-D reconstruction.

The system functional description used in this work is shown in Figure 2; correlation is used to generate the 2-D pair matches whilst the triangulation process outputs the estimated 3-D points. These processes need a priori knowledge of the camera-pair geometry to run the associated algorithms. This a priori knowledge is represented by system parameters that describe the optical and spatial relationships embedded in the captured images. The correlation algorithm searches for a given point q^l , in the left image, a similar point q^r , in the right image. This similarity evaluation is

^{*}INRIA Rhône-Alpes, 38334 St Ismier Cedex, France
{julien.morat, frederic.devernay}@inrialpes.fr

[◇]Environment Detection and Advanced Driving Assistance Systems, Electronic Department - Renault Research, France
{sebastien.cornou, javier.ibanez-guzman}@renault.com

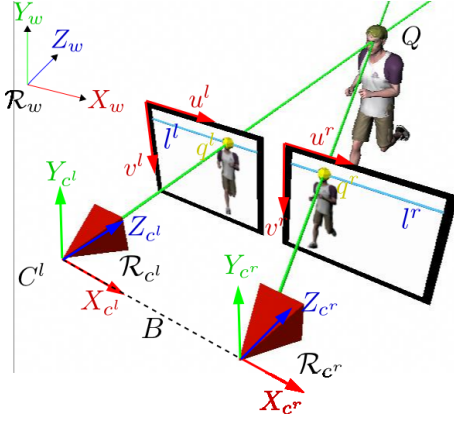


Fig. 1. Stereo-vision sensor in standard configuration, includes the optical cameras center C^l and C^r plus their reference frames $\mathcal{R}_c^l$ and $\mathcal{R}_c^r$, $\mathcal{R}_v$ is the vehicle reference frame.

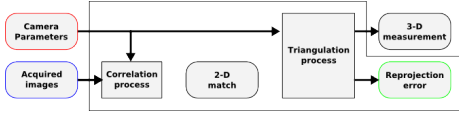


Fig. 2. Data flow diagram for the stereo-vision process.

often computationally expensive [2], [4]. Taking into account prior knowledge of the system parameters (i.e. both images are aligned), the search space of the corresponding point q^r can be reduced to a line. If the parameters are incorrect the wrong search space will be given and the correlation may fail. The triangulation process reconstructs a 3-D point Q by intersecting the lines of sight associated to the 2-D matched points q^l and q^r . The intersecting lines of sight presume knowledge of the system parameters, therefore any parameter variation results in inaccuracies and in a 3-D reconstruction error bias. It can be therefore stated that other than light conditions, estimation errors of the cameras parameters would result in accuracy errors in the 3-D reconstruction of the observed scene, or may cause the system to fail. When stereo-vision systems are mounted onboard vehicles, variations in the system parameters due to calibration errors and mechanical rigidity of the system will occur. In the following sections a new method is proposed for the evaluation of the output accuracy of a stereo-vision system for automotive applications. Section II-B details the (external and internal) camera parameters as well as the system geometry. Finally Section II-C defines the evaluation performances.

B. Camera parameters

In this paper, cameras are represented by a pinhole model with radial distortion. The camera projection model is decomposed in: Internal parameters which represent the linear projection function and the non-linear radial distortion. External parameters which represent the camera poses (position & orientation) with respect to the world reference frame. Since there are two cameras, the external parameters can be re-written in a manner to differentiate the internal geometry of the stereo-head from its external representation.

1) *Internal camera parameters:* The projection of a 3-D point expressed in the camera system coordinates onto the

sensor of a single camera is determined by its internal parameters. The linear part of the projection function can be expressed as:

$$\exists \alpha \neq 0, \begin{pmatrix} \alpha x \\ \alpha y \\ \alpha \end{pmatrix} = \begin{pmatrix} f/s_x & f/s_x \cos \theta & x_0 \\ 0 & f/s_y & y_0 \\ 0 & 0 & 1 \end{pmatrix} Q_c, \quad (1)$$

where f is the focal length in mm, s_x and s_y are the horizontal and vertical distances between pixels on the camera photo-sensor in mm, x_0 and y_0 are the principal point coordinates in pixel, θ is the angle between the u and v axes, Q_c is the 3-D point expressed in the camera reference frame $\mathcal{R}_c$, and (x, y) are the image coordinates of q . We assume $\theta = 90^\circ$, since this hypothesis is satisfied by most modern cameras. Let $f_u = f/s_x$ and $f_v = f/s_y$, therefore Eq. 1 becomes:

$$\exists \alpha \neq 0, \begin{pmatrix} \alpha x \\ \alpha y \\ \alpha \end{pmatrix} = \begin{pmatrix} f_u & 0 & x_0 \\ 0 & f_v & y_0 \\ 0 & 0 & 1 \end{pmatrix} Q_c. \quad (2)$$

The function in Eq. 2 is insufficient to model a real camera the effects of short focal lengths or other distortions introduced by the use of low cost lenses are not included. These effects can be taken into account by making the following transformation. Let $r^2 = (x - x_0)^2 + (y - y_0)^2$; the new normalized point coordinates $(\hat{x}, \hat{y})$ are defined as follows:

$$\hat{x} - x_0 = (x - x_0)(1 + \kappa_1 r^2 + \kappa_2 r^4) \quad (3)$$

$$\hat{y} - y_0 = (y - y_0)(1 + \kappa_1 r^2 + \kappa_2 r^4) \quad (4)$$

Video cameras are seldom the same, it is therefore important to take into account their own internal parameters.

2) *External camera parameters:* The external camera model describes the transform of the 3-D point Q_w , from the world reference $\mathcal{R}_w$ into a 3-D point Q_c , expressed in the camera reference $\mathcal{R}_c$, this is represented by the matrix transform Γ_w^c . The internal model can be applied only after. Γ_w^c is described by the camera pose (position and orientation) with respect to $\mathcal{R}_w$. The 6 degrees of freedom of the pose are described by the position of the origin and orientation of the reference frame, i.e. a 3-component vector for the origin and a 3-vector matrix (or a 3×3 rotation matrix) for the orientation. Interest resides in range measurements of objects in the driving environment with respect to the vehicle and not with respect to $\mathcal{R}_w$ or $\mathcal{R}_c$, therefore a new reference frame $\mathcal{R}_v$ attached to the vehicle is considered. The transformation from $\mathcal{R}_v$ to $\mathcal{R}_c$ is defined by:

$$Q_c = \Gamma_v^c(Q_v) \quad (5)$$

$$Q_v = \Gamma_w^v(Q_w). \quad (6)$$

Eq. 6 can be ignored because we are only interested in points relatively to the vehicle frame, thus the 3-D points Q_v are directly considered. To estimate range, each camera is considered, Eq. 5 is therefore applied to both cameras:

$$Q_c^l = \Gamma_v^l(Q_v) \text{ and } Q_c^r = \Gamma_v^r(Q_v). \quad (7)$$

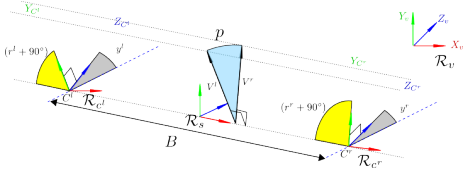


Fig. 3. Internal geometry of the stereo-vision system

3) *Internal geometry of the stereo-head*: If the camera pose is known with regard to a particular reference frame, this representation can be made with regard to any other frame, if and only if the full parameters of the transformation matrix are known. For the obstacle detection, it is necessary to know the transformation between the vehicle frame and each camera frame. The pose of every camera with respect to the vehicle should be known, however this information is difficult to obtain unless special techniques are used. If the internal geometry of the sensor from the stereo-head position in the vehicle can be separated, it would be easier to understand the sensor's behavior. For this purpose a partitioning can be made by adding a frame $\mathcal{R}_s$ attached to the sensor. The relationship between $\mathcal{R}_v$ and $\mathcal{R}_{c^i}$ ($i \in \{l, r\}$) can be written by splitting Eq. 7 as follows:

$$Q_c^l = \Gamma_s^l(\Gamma_v^s(Q_v)) \text{ and } Q_c^r = \Gamma_s^r(\Gamma_v^s(Q_v)). \quad (8)$$

where Γ_s^l (resp. Γ_s^r) is the transformation from the stereo-head reference frame $\mathcal{R}_s$ to $\mathcal{R}_{c^l}$ (resp. $\mathcal{R}_{c^r}$), and Γ_v^s is the transformation from $\mathcal{R}_v$ to $\mathcal{R}_s$.

Fixing the first camera as the stereo-head reference is a possible choice. Although, having a reference linked to the baseline seems to be more appropriate as shown in Figure 3).

The baseline B connects the optical centers C^l and C^r of the both cameras. Let Z_{c^i} , C^i and V^i be the optical axis, the optical center and vertical axis of the camera i . y^l , y^r , p^l and p^r represent the yaw and pitch angles of both cameras, with p representing the relative pitch angle between both cameras. The following equations describe the internal geometry parameters of the camera-pair:

$$\begin{aligned} b &= \|C^r - C^l\|, & B &= \frac{C^r - C^l}{\|C^r - C^l\|}, \\ y^i &= \arccos(Z_{c^i} \cdot B) - 90, & r^i &= \arccos(Y_{c^i} \cdot B) - 90 \\ V^i &= \frac{Z_{c^i} \times B}{\|Z_{c^i} \times B\|}, & p &= \arccos(V^l \cdot V^r), \end{aligned}$$

where b is the length of the baseline, B is the baseline vector, y^i the yaw angle, r the roll angle and p the relative pitch angle. All angles are expressed in degrees, $\times$ is the cross-product operator and $\cdot$ the dot product operator.

C. Performance Quantification

To evaluate in a quantified manner the performance of a stereo-vision system, a measurable set of criteria is needed. For this purpose, two criteria are defined, namely, 3-D reconstruction accuracy and reprojection error. They form a set of independent evaluation variables that permit the quantified assessment of stereo-vision systems. The first criterion represents the maximum accuracy of such system whilst the second represents the compliance of an internal

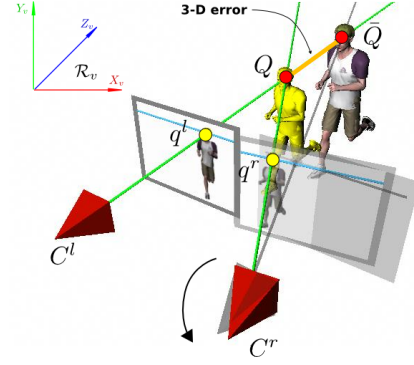


Fig. 4. 3-D reconstruction error due to yaw variation of the the right camera.

constraint needed by the algorithms. A third criterion, namely the parameter deviation, is used for the self-calibration stage. It is obtained by comparing the estimated system parameters to the actual system parameters and represents the estimation error of the self-calibration method.

These three criteria are not linearly dependent and give rise to different aspects of the system performance. The following paragraphs explain these three criteria and how to evaluate them more precisely.

1) *3-D measurement error*: The most significant criteria is the 3-D reconstruction accuracy that can be obtained by the system. This is quantified by comparing reconstructed points estimated by the stereo-vision system (measured points) with a set of points known as the ground truth. The resulting difference is defined as the 3-D measurement error along the 3-axis. It provides a quantification of the degree of accuracy of the stereo-vision system. The deviation is expressed as the Root Mean Square error (RMS) defined for each axis as:

$$\text{RMS}_K = \sqrt{\frac{1}{N} \sum_{i=0}^N (K_i - \bar{K}_i)^2}, \quad (9)$$

where K_i is one of the 3-D coordinates (X_i, Y_i, Z_i) of the i^{th} ground truth point, and $\bar{K}_i$ is the same for the i^{th} estimated point. Fig. 4 shows the type of 3-D reconstruction error generated due to changes around the yaw angle in the right hand camera alone. This will produce errors on the estimation of depth on the observed scene. In the example a positive yaw angle means that the image will appear nearer to the vehicle than what it is in reality (Q and $\bar{Q}$).

2) *Reprojection error*: A low reprojection error is required by the internal process of the stereo-vision system. It represents the compliance of a geometric constraint used in the correlation process. It can provide information on the possible failure of the correlation process and then the subsequent entire system failure. That is, left and right images in a camera pair are acquired at the same time, the cameras are roughly aligned and their geometrical relationship is known. This configuration induces a relationship between both images, called the epipolar constraint. If this constraint is not consistently verified, either the correlation may fail or the epipolar constraint has to be relaxed. For the latter, the triangulation process will be unsuccessful, thus the lines of sight of the two matching points will not intersect, and no

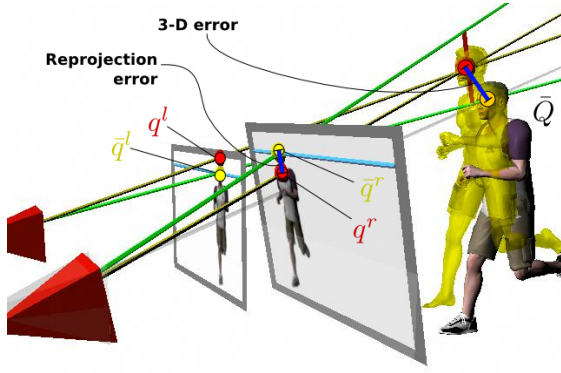


Fig. 5. 2-D reprojection error due to pitch variation of the the right camera.

reconstructed point can be determined. In practice, the epipolar constraint is not verified, and triangulation algorithms return the optimal 3-D point for the given measurements. Consequently, the image projections of the reconstructed points do not match the initial measure. The reprojection error can be expressed in pixels as the distance between the reprojected 2-D point and the 2-D measure from which it is issued. For this purpose the Root Mean Square error is computed for the y coordinates of the $2N$ points (left and right images) as follows:

$$\text{RMS}_k = \sqrt{\frac{1}{2N} \sum_{j \in \{l, r\}} \sum_{i=1}^N (\bar{k}_i^j - k_i^j)^2}, \text{ with } k \in \{x, y\}. \quad (10)$$

That is, the reprojection error means that the epipolar constraint is not verified. In fronto-parallel case, the epipolar lines are horizontal, a pure reprojection error along the x -axis will give a point belonging to the epipolar line. That is a contradiction for a point that should break the epipolar constraint. Consequently the value for the RMS_x is ignored. Fig. 5 shows the 2-D reprojection error that occurs due to variations around the pitch angle of the right hand camera in the stereo pair. This type of error might produce the wrong estimations of the obstacle height and thus the wrong decision could be taken. Since the reprojection error is induced by inconsistent epipolar geometry, it will inform whether or not this constraint has been satisfied, though it does not provide a direct measure of the sensor accuracy. But, since the triangulation process relies on this constraint, a high reprojection error informs about the high probability of failure of that process.

3) *Deviation of camera parameters:* This is defined as the difference between the true and estimated parameters. Therefore, parameters estimated after the self-calibration processes are available. Whilst the deviation of the estimations with respect to true values does not inform of the 3-D reconstruction accuracy, it provides feedback that helps to have a better understanding of error propagation through the calibration process. In order to facilitate the visualization of external parameters, we present them with respect to the frame $\mathcal{R}_s$ (c.f. II-B.3) of the stereo-vision system. For the same reasons, only the focal length will be presented for the internal parameters.

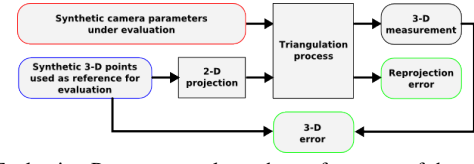


Fig. 6. Evaluation Process to evaluate the performance of the stereo-vision system due to parameter deviation.

III. SYSTEM EVALUATION DURING THE SYSTEM LIFE-CYCLE

Calibration is a very important process in stereo-vision and should be performed in the factory floor. Once the vehicle is in operation, it will suffer from strong mechanical vibrations, changes in temperature or small shocks and thus the spatial relationship and other camera parameters will vary. Changes in the spatial relationship between cameras will results in changes on the calibration properties of the sensor and this will become invalid. For example, depth estimation errors would occur. Therefore for a vehicle manufacturer, it is important to determine the expected errors that might exist for a given parameter deviation. For this purpose a method is proposed to observe the potential effects of such changes. The fundamental question here is: What is the expected error for a given parameter deviation?

A. Method to Evaluate the Consequences of Parameter Changes

In order to evaluate the stereo-vision system, the system inputs and outputs are generated. First, this way we can control the data as well as the noise which can be added to the inputs. Second, because the exact ground truth is known, the estimations obtained by the algorithms can be compared. Whilst, it could have been possible to use synthetic images with the algorithms, it was preferred to use numerical data only in order to overcome search problems.

The three synthetic inputs used are the true parameters of the stereo-vision system, the true parameters plus a known deviation, and a set of known 3-D points distributed in the the near and far zones. By applying the system process over both sets of data it is possible to obtain the pairs of 2-D true points, and the pairs of measured 2-D points corrupted with noise. These are then used to reconstruct the 3-D corrupted points. The measurement error is obtained by comparing the corrupted 3-D points with the original ones. The reprojection error is obtained by comparing the measured 2-D points obtained by the reprojection of the rebuilt 3-D points. The block diagram in Figure 6 illustrates the principle as a modification to the original stereo-vision process described in the previous section. Because the method is based on runs, each possible configuration could be tested. However the number of possible configurations is too high for all to be tested. For the simulations, a single parameter variation at a time is taken into account. Consequently configurations having only the variation of a single parameter are considered. Assuming that the variations are small and that the function is locally linear, a combination of multiple simultaneous variations is less or equal to the sum of the errors obtained from the same variations are considered independent. Moreover, considering

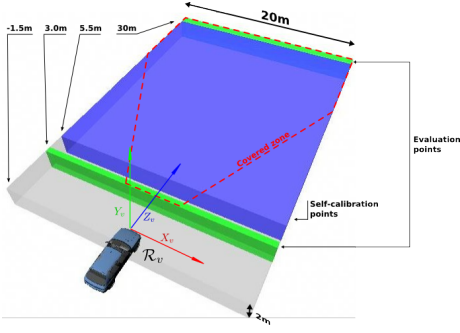


Fig. 7. Experimental setup for evaluation purposes. Reference points for self-calibration are generated for $5.5m \leq Z_v \leq 30m$.

one parameter at a time also allows us to limit the evaluation to parameters of a single camera, others can be obtained by symmetry. That is the results obtained for the right camera can be extended to the left camera. Other specific parameter variations could be considered, for example, the simultaneous variation of the focal-lengths of both cameras due for example to temperature changes affecting both cameras, but these are considered as special cases.

B. Experimental Setup

The numerical results obtained depend on the camera parameters (true and deviated values), but also the 3-D points used to sample errors. In the results presented, the input and evaluation values have been fixed to representative ones. The pair of stereo-vision synthetic cameras used is located behind a vehicle windscreen. That is at its middle top part, around the rear view mirror. The baseline (B) of the cameras is 40 cm. The cameras have a focal length of 6 mm for a $1/3''$ imaging sensor, with a resolution of 640×480 pixels ($f = 800$ pix), and radial distortion coefficients set to zero. The principal point for each camera is placed in the middle of the image. The vehicle reference frame $\mathcal{R}_v$ is defined as being on the road plan along the vehicle longitudinal axis and the origin of the vehicle hood as shown in Fig. 7. The cameras are therefore defined as being at $(\mp 0.2m, 1.4m, -1.5m)$ in $\mathcal{R}_v$.

C. Results

Whilst different amplitudes for the distortions were examined in the experimentation phase, in this paper due to space constraints only deviations by 0.5° in yaw and pitch angles as well as a 0.5% deviation in focal length are used for illustration purposes. These values are extreme with regard to the measurements made on a test vehicle. The errors generated by these three variations are shown in Fig. 8. A summary of the numerical results of the referred variations are shown in Table I. They demonstrate how a 0.5° deviation on the yaw angle has a strong impact on 3-D measurement accuracy, while the reprojection error remains insignificant. By contrast, a small deviation of pitch angle has a strong effect on reprojection error while 3-D measurement accuracy remains acceptable. In the first case, the system returns an inaccurate measure, whilst in the second the correlation process may totally fail leading to the entire system failure. The table also shows that measurement imprecision is more

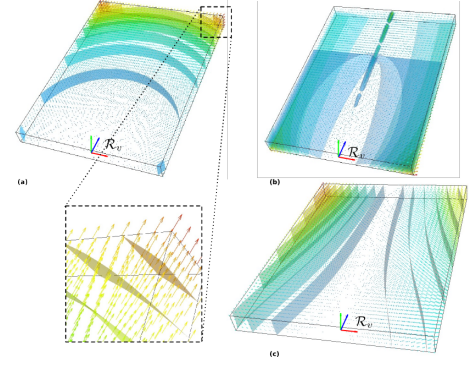


Fig. 8. 3-D Reconstruction error for (a) $y^r + 0.5^\circ$, (b) $p^r + 0.5^\circ$ and (c) $f^r + 0.5\%$. Vectors represent the deviation and trips are iso-surfaces of the norm vector function.

TABLE I
Results of Pitch, Yaw and Focal Length deviations the perception zones of interest of the stereo-vision system

deviation	$pitch^r + 0.5^\circ$		$yaw^r + 0.5^\circ$		$focal^r + 0.5\%$	
sample dist. (m)	3	30	3	30	3	30
RMS _X (cm)	1.03	10.2	22	570	5	265
RMS _Y (cm)	2.6	14.4	8.7	31.3	1.5	11.2
RMS _Z (cm)	2.7	20.8	74.6	1480	13	528
RMS _y (pixel)	3.54	3.49	0.11	0.02	0.2	0.04

important in range (depth). The worst case for the loss of precision occurs when there is a variation around the yaw angle, in particular for points which are located far from the vehicle. It should be noted that this miss-calibration does not lead to a violation of the epipolar constraint which would result in a large reprojection error. A total of 4800 3-D points were used located at the nearest and furthest zones, that are detailed as follows. The maximum errors are for each lane entry are represented in bold. As referred previously, the results depend very much in the 3-D samples used in the evaluation of the error. Fig. 8 indicates that the measured error of a stereo vision system depends on the relative position of the measured point. For example, the uncertainty is generally bigger for far away points. It has been necessary therefore to segment the perceived space of $20 \times 27m$ into various zones of interests. Fig. 7 present the two zones used in the evaluations: the closest one in green (at 3m in front of the vehicle) and the furthest also in green (at 30m from the vehicle). These zones are the areas of interest sampled by the grids of 3-D points. As it is emphasize in table I, small deviation of yaw has a strong impact on 3-D measurement accuracy, while reprojection error remains non-significant. At the opposite, a small deviation of pitch has a strong effect on reprojection error while 3-D measurement accuracy remains acceptable. In the first case, the system returns an inaccurate measure, whilst in the second the correlation process may totally fail leading to the entire system failure.

IV. SELF-CALIBRATION PROCESS

It has been shown in Section III that when the camera parameters deviate, the system returns corrupted measurements or even breaks down. For systems to be mounted onboard mass market vehicles the only way will be to recalibrate the cameras, most likely based on the corrupted measurements

rather than using conventional calibration targets due to the logistics involved [6]. Therefore, a stereo-vision system equipping a series of vehicle must be able to re-calibrate itself, or at least to self-diagnose system failure in order to ensure its usability. One approach is to apply a bundle adjustment to the stereo-vision rig in order to effect some self-calibration. In this paper, the focus is on evaluation method that identifies potential sources of inaccuracy occurring in the self-calibration process. This is verified through a set of simulations similar to those presented in the previous section.

A. Self-calibration method applied to case of an onboard stereo-vision system

A recent focus of research has been in the area of self-calibration. The rationale being that it is possible to calibrate cameras from the data available during use rather than having to rely on special purpose calibration data. A versatile and accurate self-calibration method is the bundle adjustment. Based on a non-linear minimization, it produces joint parameter estimates and 3-D reconstruction. Basically, the minimized cost function relies on reprojection error criteria. Generic methods for bundle adjustment [7], [8] are not well-adapted to the particular case of automotive stereo-vision system, notably because the 2 points of view are close to each other, and the scene configuration (quasi-planar road) is a degenerate case for these methods. To make the process efficient, additional constraints should be considered. [9] chose to add some prior knowledge by placing visible targets fixed to the car. We do not prefer adding features to the car but to use properties of the mechanical behavior of the cameras. The cost function becomes:

$$\alpha_{2D} \sum_{i=1}^m D(\tilde{q}_i^c, q_i^c)^2 + \sum_{c \in \{l, r\}} \sum_{j=1}^n \alpha_j (\hat{p}_j^c - p_j^c)^2 + \alpha_b (\hat{b} - b)^2 \quad (11)$$

with m equal the number of measured pair of points, n equal the number of estimated parameters, $\tilde{q}_i^c$ is the i^{th} point projected in the image c , q_i^c the measure of the same point, D the Euclidean distance, $\hat{p}_j^c$ and p_j^c the initial and estimated parameters for camera c , $\hat{b}$ and b the initial and estimated baseline length, and finally α the normalization factors. The camera model must be specified with a minimum number of parameters p_j . Most of internal parameters are very difficult to recover with the bundle adjustment. This is the reason why we evaluate only the focal length. All external parameters are estimated in the vehicle reference frame $\mathcal{R}_v$.

The reprojection error was introduced based on the assumption of a Gaussian error distribution of the reprojected points. This assumption is valid when all the matches (correspondences) are correct. However, in our application, we also need to consider the error caused by false matches. If we consider these outliers, the error distribution is no longer Gaussian. This error can be handled in the optimization procedure by applying robust estimation techniques on the term $\sum_{i=1}^m D(\tilde{q}_i^c, q_i^c)^2$ of Eq. 11, such as M-Estimator [10], [11].

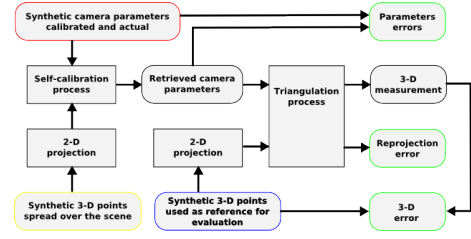


Fig. 9. Evaluation of the system performances of self-calibration method facing parameters deviations.

B. Evaluation method extended to systems including self-calibration process

Stereo-vision systems with self-calibration capabilities are in general more complex than those with standard fixed parameter configurations. Therefore it is important to estimate the benefits brought by self-calibration methods in particular if these are to be incorporated as part of automotive solutions in order to ensure the proper operation of applications based on stereo vision systems. The proposed evaluation method is summarized in block diagram form in fig. 9. It is formulated under the same scheme used for the evaluation of standard stereo vision systems discussed earlier. First, a deviation is added to particular parameters of the stereo camera pair, for example, the yaw angle of the stereo-vision system can be modified to emulate the loss of accuracy either during assembly or following a vehicle collision. Then, a first set of points (the blue points in fig. 7) is used to perform the self-calibration. These points are projected in each image according to the stereo-vision system deviated parameters (i.e. as they are). The obtained projections are used to perform the self-calibration. The resulting system parameters are then considered as the current state. Once the parameters have been updated, another set of points (the green points in fig. 7) is used to estimate the current system accuracy as described in Sec. III. The reprojection and 3-D error, the distance between the initial camera parameters and the estimated camera parameters are considered. This framework can be used to estimate the maximum error resulting of deviations on a set of stereo-vision parameters. To analyze further the benefits of the self-calibration process to coupled deviations of the parameters, the proposed method can be applied as sets of deviations generated as a random process on all the system variables that are susceptible to change. In that case, sets of random parameters can be generated to allow a statistical analysis of the benefits induced by the self-calibration process according to a large number of camera parameter deviations. In such a case, it should be necessary to perform a design of experiment analysis.

C. Experiment and Results

The same experimental setup as for the calibration process is used. The results of the calibration process have shown the effects on the system performance on variations along the yaw and pitch angles as well as focal length.

In this experiment, it is assumed that these changes can be detected and consequently the self-calibration process has been applied. A very important issue is thus to determine

TABLE II

Evaluation of the bundle adjustment technique for the variations presented in Fig. I. 800 points the blue volume of Fig. 7 were used to self-calibrate the system with $\sigma_{2D} = 0.1\text{pix}$; evaluation was made with 4800 3-D reference points at 30m distant (green zone on Fig. 7).

Deviation	$pitch^r + 0.5^\circ$	$yaw^r + 0.5^\circ$	$focal^r + 0.5\%$
RMS_X (cm)	~ 0	209.9	56.5
RMS_Y (cm)	~ 0	27.6	11
RMS_Z (cm)	~ 0	898	184
RMS_y (pix)	~ 0	3.5	3.4
$f_u^l \times f_v^l$ (pix)	800×800	800×800	800×799.7
$f_u^r \times f_v^r$ (pix)	800×799.9	799.9×799.9	799.9×800.2
b (mm)	399	399	40
p ($^\circ$)	0.50	~ 0	~ 0
$r^l r^r$ ($^\circ$)	0.06 0.06	-0.08 -0.08	0.12 0.12
$y^l y^r$ ($^\circ$)	0.95 0.77	-0.22 -0.10	6.24 6.20

how the effects of these changes have been corrected or reduced once the self-calibration process has been applied. Application of the formulated performance criteria gives a very good insight into this issue as demonstrated in Table II. It summarizes the results of the evaluation. This has been applied on 800 points within the blue volume in fig. 7. These were used to calibrate the system with $\sigma_{2D} = 0.1$ pixel. For furthest distance at 30m, the furthest green volume in fig. 7, 4800 points were used. The results summarized in Table II indicate that the self-calibration method does not respond equally. It changes according to the variations. For example the error due to an angular change around the yaw axis was equal a RMS error of 1480, after an application of the self-calibration process, the RMS error became 898. Whilst there is an overall improvement to 66%, there is still a substantial error. It should be remarked that a 0.5° variation around the yaw angle might represent a large magnitude, and that the distance at which the error is calculated, it is the furthest. The self-calibration results around the variations in the pitch angle are very good, the 3-D and the reprojection errors are completely corrected. Self-calibration also corrects the variation effects on the focal length. Table II shows also that a RMS depth estimation error of 528 for a 0.5% variation in focal length becomes equal to an RMS of 184, an improvement of 34.84%. It is clear that the variations in yaw are corrected but not sufficiently whilst variations in the pitch angle are completely recalibrated. This can be understood as self-calibration seeks parameters that satisfy the epipolar constraint whilst a yaw variation does not affect very much this criterion. For a designer of automotive stereo vision perception systems, this is very important because fixtures for example have to privilege constraints around the yaw angle rather than pitch. The table II shows that the presented method works better against pitch deviation than against yaw deviation. Since the method minimize the reprojection error and a pitch deviation leads to high reprojection error

V. CONCLUSION

The quantitative evaluation of stereo-vision systems on-board of mass market vehicles has been presented in this paper. The fundamental evaluation is made by considering the geometric relationships that exist when a camera-pair

is used for stereo-vision purposes. The method takes into account the system configuration and variations which might occur during the vehicle working life. For this purpose two issues were examined, the effects of variations in the physical properties of the system and the effects on the range estimation process, as well as the performance of self-calibration processes as applied to these variations. Both results enable system designers to assess what is expected when confronted with variations but also to assess the performance of self-calibration methods. As a result, design rules can be generated to constraint mechanically the variations along or around certain directions like in the case of the yaw angle. The simulations have allowed the exploration of a large number of possible configurations. This evaluation is made following two criteria, the first looks into the precision of the measurement and the second is with respect to the epipolar constraint necessary for stereo-vision algorithms to work in a good manner. The results have shown that the stability of the yaw and pitch angles are very important as the precision of the measurements depends very much on them. It was also possible to infer that self-calibration methods can correct successfully variations around the pitch angle whilst for corrections around the yaw angle, this is more difficult. The results are to be incorporated into the design criteria to be used for the deployment of stereo-vision systems in mass market vehicles.

REFERENCES

- [1] V. Srinivasan, "A vision for supporting autonomous navigation in urban environments," in *IEEE computer*, pp. 68–77, 2006.
- [2] O. Faugeras, B. Hotz, H. Mathieu, T. Viéville, Z. Zhang, P. Fua, E. Thérion, L. Moll, G. Berry, J. Vuillemin, P. Bertin, and C. Proy, "Real time correlation based stereo: algorithm implementations and applications," Tech. Rep. RR-2013, INRIA, 1993.
- [3] U. Franke and S. Heinrich, "Fast obstacle detection for urban traffic situations," in *Intelligent Transportation Systems, IEEE Transactions*, vol. 3, pp. 173–181, Sept. 2002.
- [4] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, no. 1-3, pp. 7–42, 2002.
- [5] W. van der Mark and D. M. Gavrila, "Real-time dense stereo for intelligent vehicles," in *Proceedings of Conference Intelligent Transportation Systems, IEEE Transactions*, pp. 38–50, 2006.
- [6] H. Beyer, "Accurate calibration of CCD cameras," in *Proceedings of Conference on Computer Vision and Pattern Recognition CVPR92*, (Urbana-Champaign, USA), pp. 96–101, 1992.
- [7] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment — a modern synthesis," in *Vision Algorithms: Theory and Practice* (B. Triggs, A. Zisserman, and R. Szeliski, eds.), no. 1883 in LNCS, (Corfu, Greece), pp. 298–373, Springer-Verlag, Sept. 1999.
- [8] M. Lourakis and A. Argyros, "Is levenberg-marquardt the most efficient optimization algorithm for implementing bundle adjustment?," in *ICCV 2005. Tenth IEEE International Conference on Computer Vision*, vol. 2, pp. 1526–1531, 2005.
- [9] M. Bertozzi, A. Broggi, and A. Fascioli, "Self-calibration of a stereo vision system for automotive applications," in *Proc. IEEE Intl. Conf. on Robotics and Automation*, vol. 4, (Seoul, Korea), pp. 3698–3703, May 2001.
- [10] J. W. Tukey, *Exploratory Data Analysis*. Addison-Wesley, 1977.
- [11] P. J. Huber, *Robust Statistics*. John Wiley and Sons, 1981.